An Optimal Rewiring Strategy for Cooperative Multiagent Social Learning

Hongyao Tang,[1] Jianye Hao,[1] Li Wang,[1] Tim Baarslag,[2] Zan Wang[1]
[1]College of Intelligence and Computing, Tianjin University, China
[2]Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
{bluecontra, jianye.hao, wangli, wangzan}@tju.edu.cn, T.Baarslag@cwi.nl
+86 15620947182, Yaguan Road 135, Jinnan Distinct, Tianjin, China

Abstract

Multiagent coordination in cooperative multiagent systems (MASs) has been widely studied in both fixed-agent repeated interaction setting and static social learning framework. However, two aspects of dynamics in real-world MASs are currently missing. First, the network topologies can dynamically change during the course of interaction. Second, the interaction utilities between each pair of agents may not be identical and not known as a prior. Both issues mentioned above increase the difficulty of coordination. In this paper, we consider the multiagent social learning in a dynamic environment in which agents can alter their connections and interact with randomly chosen neighbors with unknown utilities beforehand. We propose an optimal rewiring strategy to select most beneficial peers to maximize the accumulated payoffs in long-run interactions. We empirically demonstrate the effects of our approach in large-scale MASs.

Introduction

Multiagent coordination in cooperative multiagent systems (MASs) is a significant and widely studied problem. It requires agents to have the capability of coordinating with others effectively towards desirable outcomes. Until now, a lot of works have studied the multiagent coordination problems in cooperative MASs (Claus and Boutilier 1998). One class of research is multiagent social learning (Sen and Airiau 2007), which study the multiagent coordination problem among a population of cooperative agents with sparse and local interactions (Hao and Leung 2013).

Most existing works under the social learning framework assume that agents are located in a static network. However, two important aspects of dynamics in real-world MASs are currently missing. First, the cooperative games for agent pairs might be different due to the difference of agents' preferences and the contexts they are situated in. Second, the network topologies can be dynamic, i.e. agent's interacting partners are not fixed and may change frequently. Therefore, in this paper, we study the multiagent coordination problem in cooperative MASs with taking above two aspects into consideration. We consider a dynamic environment where agents can alter their connections through rewiring autonomously, and propose an optimal rewiring approach for agents to select most beneficial peers among all reachable peers to maximize the accumulative payoff during the long-run interactions.

Problem Description

We consider a population of agents N, in which each agent i can only play with its *reachable peers*, defined as $\{O_i \cup \bar{O}_i\}$. Agent i can only interact with its neighborhood O_i through the connections, and also has a probability φ to be able to establish a new connection to a potential agent $j \in \bar{O}_i$ with cost c_j^i through rewiring. For each rewiring, an old connection should be broken before establishing a new one to model agents' limited communication ability in practice.

We model the strategic interaction among each pair of agents as cooperative games. A general form of two-action cooperative games between agent i and j is denoted as $G_i^j = [u_a, \alpha, \alpha, u_b]$, where u_a (or u_b) is the payoff when agent i and j both choose action a (or b) and α ($\leq u_a(u_b)$) is the payoff for mis-coordination. To model the uncertainty and diversity of agents' utility functions, the coordination payoff u_a (or u_b) is sampled from a stochastic variable x_a (or x_b) following cumulative probability distribution $F_a(x)$ (or $F_b(x)$). In addition, $F_a(x)$ (or $F_b(x)$) is unique for each game. The value of u_a (or u_b) is unknown before interaction and is revealed when the corresponding outcome is reached once. Each agent can observe the actions of its interaction neighbor at the end of each interaction.

Social Learning Framework

The overall social learning framework is shown in Algorithm 1. During each round, agent $i \in N$ goes through the rewiring phase first and then the interaction phase.

Estimation of Expected Interaction Payoff The expected payoffs of agent i interacting with j according to an known or unknown payoff matrix G_i^j, i.e., v_i^j or x_i^j, are evaluated respectively as follows:

$$v_i^j = \max_{m \in A_i} p_i^j(m) u_m + \left(1 - p_i^j(m)\right) \alpha, \tag{1}$$

$$x_i^j = \max_{m \in A_i} p_i^j(m) x_m + \left(1 - p_i^j(m)\right) \alpha. \tag{2}$$

Agent j's policy p_i^j can be estimated from historical actions.

Algorithm 1 Overall interaction protocol for agent $i \in N$.

for a number of interaction rounds **do**
  if random variable $p \leq \varphi$ **then**
    Perform rewiring action (including **NOOP**).
  end if
  Play game G_i^j with randomly chosen player $j \in O_i$.
  Obtain payoff and update its policy.
  Update neighbor j's action model.
end for

Algorithm 2 K-Sight rewiring strategy for agent i with sight K in each rewiring phase.

for each $j \in O_i$ **do**
  Compute expected payoff v_i^j (Eq.1).
end for
Obtain the interaction baseline $y_i = \min_{j \in O_i} v_i^j$.
for each $w \in \bar{O}_i$ **do**
  Compute benefit index Λ_i^w for agent w (Eq.3).
end for
Choose agent $t = \arg\max_{w \in \bar{O}_i} (K\Lambda_i^w - c_i^w)$.
if $K\Lambda_i^w - c_i^w \geq 0$ **then**
  Rewire agent t and break the worst connection.
end if

An Optimal Rewiring Strategy Each agent's situated environments are continuously changing due to rewiring and thus we model it as an Markov Decision Process. We propose K-sight Rewiring Strategy (Algorithm 2). This is inspired from Pandora's Rule (Weitzman 1979) and Negotiation Problem (Baarslag and Gerding 2015), and the optimality can be similarly proved as did in above mentioned works. An index Λ_i^j is calculated to captures the benefits of rewiring unknown peers j:

$$\Lambda_i^j = \int_{-\infty}^{\infty} y_i' \cdot \mathrm{d}F_i^j(x) - y_i, \tag{3}$$

where y_i and $y_i\prime$ are the minimum of $\{v_i^j\}_{j \in O_i}$ before and after rewiring. $F_i^j(x)$ is the distribution of x_i^j (Eq.2).

Interaction Strategies We consider three strategies in interaction phase, i.e., Fictitious play (FP), Joint-Action Learner (JAL) and Joint-Action WoLF-PHC (JA-WoLF).

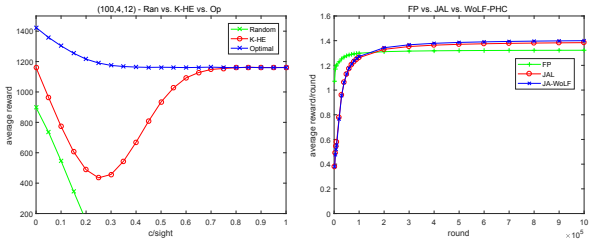
Experimental Evaluations

We compare our approach (Optimal) with two benchmark strategies, i.e., Random (Ran) and K-sight Highest Expect (K-HE) that rewires the agent with the highest positive value of K-round expected payoff minus the cost.

First, we evaluate the performance of each rewiring strategy in environments of different scales. The average accumulated payoff over 1000 rounds of each agent are shown in Table 1. We can observe that our optimal rewiring strategy outperforms benchmark strategies in terms of average, best and worst cases across all settings. Second, in Figure 1(a) we evaluate our approach under the settings with the rewiring cost c varying in the range of $[0.0, 200.0]$ and the fixed $K = 200$. The results show that our approach significantly outperforms others across almost all c/K settings. We use FP as interaction strategy for Table 1 and Figure 1(a).

Table 1: The performance of rewiring strategies in different topologies with $c = 20.0$, $\varphi = 0.01$. x, y, z denote the initial size of agents, neighborhood and reachable peers.

No.	(x, y, z)	Rew_Stg	Avg.	Max.	Min.
1	$(100, 4, 12)$	Random	763	1383	283
2	$(100, 4, 12)$	K-HE	1003	1979	306
3	$(100, 4, 12)$	**Optimal**	**1372**	**2437**	**665**
4	$(500, 4, 16)$	Random	694	1434	169
5	$(500, 4, 16)$	K-HE	993	2173	211
6	$(500, 4, 16)$	**Optimal**	**1373**	**2735**	**484**
7	$(1000, 8, 16)$	Random	740	1375	217
8	$(1000, 8, 16)$	K-HE	1018	1801	314
9	$(1000, 8, 16)$	**Optimal**	**1170**	**1810**	**567**



(a) (b)

Figure 1: Performance comparison for (a) different rewiring strategies and (b) different interaction strategies.

Moreover, we analyze the performance of three interaction strategies, i.e., FP, JAL, JA-WoLF. The results are shown in Figure 1(b) with regard to the average single-round interaction payoff of each agent. We can observe that FP strategy can fast reach a good payoff level while JAL and JA-WoLF outperform FP in the long term due to their better convergence on optimal Nash equilibrium.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 61702362) and is part of the Veni research programme (No. 639.021.751) financed by the Netherlands Organisation for Scientific Research (NWO).

References

Baarslag, T., and Gerding, E. H. 2015. Optimal incremental preference elicitation during negotiation. In *IJCAI*, 3–9.

Claus, C., and Boutilier, C. 1998. The dynamics of reinforcement learning in cooperative multiagent systems. In *AAAI*, 746–752.

Hao, J., and Leung, H. 2013. The dynamics of reinforcement social learning in cooperative multiagent systems. In *IJCAI*, 184–190.

Sen, S., and Airiau, S. 2007. Emergence of norms through social learning. In *IJCAI*, 1507–1512.

Weitzman, M. L. 1979. Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society* 641–654.